<table>
<thead>
<tr>
<th>Title</th>
<th>Bilingual First Language Development of Two Japanese-English Bilingual Siblings: Literature Review on Possible Dependent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>TAURA, Hideyuki; TAURA, Amanda</td>
</tr>
<tr>
<td>Editor(s)</td>
<td></td>
</tr>
<tr>
<td>Citation</td>
<td>言語文化学研究 言語情報編 2010, 5, p.21-46</td>
</tr>
<tr>
<td>Issue Date</td>
<td>2010-03-31</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10466/11456">http://hdl.handle.net/10466/11456</a></td>
</tr>
<tr>
<td>Rights</td>
<td></td>
</tr>
</tbody>
</table>

http://repository.osakafu-u.ac.jp/dspace/
Bilingual First Language Development of Two Japanese-English Bilingual Siblings: Literature Review on Possible Dependent Variables

TAURA, Hideyuki & TAURA, Amanda

Journal of Language and Culture
Language and Information
Vol. 5 (2010)
Department of Language and Culture
School of Humanities and Social Sciences
Osaka Prefecture University

言語文化学研究（言語情報編）
第 3号抜刷
大阪府立大学人間社会学部 言語文化学科
Bilingual First Language Development of Two Japanese-English Bilingual Siblings: Literature Review on Possible Dependent Variables

TAURA, Hideyuki & TAURA, Amanda

1. Introduction

Research into Japanese-English bilinguals ($N=2$) who have been mostly brought up in Japan (apart from 2 years in Australian school settings) to disclose developmental stages of one of their first languages, English, is the primary goal in our bigger project. The longitudinal data collected to date from two bilingual siblings include (1) spoken data from ages 8;06 to 19;10 and written data from 12;02 to 18;05 years with the elder brother and (2) spoken data from ages 4;09-18;00 and written data from 8;05-18;00 years with the younger sister. We intend to analyze the bilingual data from a developmental perspective in comparison to findings from English monolinguals' developmental studies in the four skill areas of speaking, writing, listening, and reading. The data are examined at such linguistic levels as vocabulary, syntax, and morphology for fluency, accuracy, and complexity analysis.

The development of the L2 is claimed to be similar to that of one's L1 as long as enough exposure is provided in both languages (Genesee, 2004a; Weber-Fox and Neville, 2001), which applies to our discussion in this paper on bilingual first language acquisition (BiL1, henceforth). Genesee (2004a: 279) further states that "key milestones in phonological, lexical, syntactic, and pragmatic development occur within the same age range for bilingual children as for monolingual children." With regard to one's L2 acquisition, however, the linguistic distance between the two languages has to be taken into account as Hakuta (2000) argues on how difficult it is for a Japanese native speaker to acquire English articles because no such
grammatical morphemes exist in Japanese. Would Japanese-English bilingual children acquiring two languages simultaneously from birth follow the same acquisitional course taken by English monolingual children? If the path of English acquisition is similar in childhood, the query arises as to whether language development during adolescence and adulthood is also the same since "language is presumed to have developed on a solid foundation established during infancy and early childhood" (Nippold, 2006: 11). If language development in bilinguals during adolescence and adulthood does correspond to their monolingual counterparts, would it be of a gradual and subtle nature? In trying to find the answers to these questions, the present researchers propose tasks (1) that are challenging such as abstract noun definition (McGhee-Bidlack, 1991), (2) that examine low frequency grammatical structures and intersentential elements in both spoken and written contexts (Nelson, 1988; Nippold, Schwarz, and Undlin, 1992; Scott, 1988b), (3) that require children to explain idioms and proverbs (Nippold and Rudzinski, 1993), and (4) that make it possible to observe how bilingual children change the style of their speech to communicate with different listeners for varied purposes (Nippold, 1995; Selman, Beardslee, Schultz, Krupa, and Podorefsky, 1986).

As the first step, the present study attempts to narrow down on a set of variables to use in the second stage onwards, which are best suited for our purpose. Relevant developmental studies on English as a first language both in monolinguals and bilinguals, though with the latter, cases are scarce in the literature, are reviewed.

Research design (i.e., cross-sectional vs. longitudinal) is briefly discussed first, followed by the literature review which is the main section of this paper. It is hoped that we can end the current attempt with a set of variables that can be used in our study and also throw a light on previous developmental studies.
2. Cross-sectional vs. longitudinal design

Both cross-sectional and longitudinal studies should ideally be used to support and complement the findings or shortcomings in any linguistic inquiry. However, due to the cost and time involved in longitudinal studies, most tend to employ a cross-sectional or quasi-longitudinal design. In other words, a number of children represent each developmental stages, for example, pre-school, primary school, high school, or university. Needless to say, the data collected in this way must be handled with caution as de Bot (2005) claims that the statistical means of a group of people do not really represent one particular participant's characteristics. At the same time, longitudinal data are also not flawless though a deep insight into the mechanism of language may be gained through rich data. The findings drawn from one or even a few cases can simply not be generalized over a bigger population. Considering the relative abundance of cross-sectional studies on bilingual language acquisition (i.e., de Groot and Kroll, 1997; Kecskes and Albertazzi, 2007; Montrul, 2008) over longitudinal or case studies (i.e., De Houwer, 1990; Lanza, 2004; Leopold, 1939, 1947, 1949a, 1949b, 1954; Swain and Wesche, 1975; Vihman, 1985; Volterra and Taeschner, 1978), the current study has adopted a longitudinal design. This does not mean, however, that cross-sectional studies have been eliminated in our quest to identify a set of variables best suited to our study. Our literature review, therefore, includes both cross-sectional and longitudinal studies as long as they are concerned with English language acquisition development either in monolinguals or bilinguals.

3. Literature review

The literature review in this section is as follows: the acquisition of lexicon, metaphors and similes, morphology, syntax, speaking (conversation and narration), and writing. This categorization of language aspects is partly due to the findings from Nippold's (2006) rigorous literature review on first
language development. Nippold found that L1 development continues throughout the school-age and adolescent years and well into adulthood in vocabulary (lexicon), morphology, and syntax, despite the findings that most five-year-olds can (1) recognize over 10,000 words (Clark, 2003), (2) produce syntactically complex sentences with relative clauses and infinitives (Paul, 1981), and (3) speak intelligibly more than 90% of the time (Clark, 2003), results which appear to lend support to the view that L1 acquisition is virtually complete at age five. The inclusion of literacy skill development (understanding literate meanings of metaphors and similes, story-telling, and writing) is explained by the fact that in the third or fourth grade (ages 8 to 10), children begin using reading skills to learn advanced vocabulary and complex syntax, which enables them to acquire a great deal of linguistic and world knowledge independently and pursue personal interests more readily (Reed, 1986). As a result, children become increasingly individualistic in their language development during the school-age and adolescent years (Gallatin, 1975; Nippold, 1995). This is one of the areas where our two siblings differ from each other - the elder brother loves reading in both English and Japanese while the younger sister does not, which leads to one of our research questions - individual differences in bilingual first language acquisition between siblings.

3.1 Lexicon

Children are believed to acquire 2,000 to 3,000 words each year during their school years, especially when they begin reading independently from age 9 or 10 and the written language becomes a significant source of lexical learning, until they graduate from high school with a repertoire of at least 40,000 words (Nagy and Scott, 2000; White, Power, and White, 1989). In spite of children's spontaneous use of a wide variety of words, research shows (e.g. Nippold, 2006; Miller and Gildea, 1987; Perera, 1986) that a full understanding of certain types of words may not be reached till late
adolescence as the subtle meanings of literate words are acquired gradually during the school-age years, adolescence, and early adulthood. In such studies, literate words mean (1) polysemous words, (2) double-function words, (3) adverbs (of likelihood and magnitude), (4) abstract words, (5) metalinguistic and metacognitive verbs, and (6) factive and nonfactive verbs.

A brief summary of the research that looked into the developmental acquisition of these six types of literate words is provided here. First, polysemous words such as 'up', 'above', and 'low' are used by preschool-age children but their secondary meanings are acquired a lot later, sometimes in adolescence (Durkin et al., 1985). Comprehension of these words showed improvement as a function of increasing grade level in primary school (Mason et al., 1979).

Secondly, children understand the physical meaning of double-function words like 'cold', 'warm', 'sweet', 'hard', 'soft', 'bright', 'deep', and 'crooked' several years before they understand the psychological meanings. Asch and Nerlove (1960), for example, examined 'sweet' and 'hard' and found that children at ages 5 and 6 years begin to understand the psychological meanings.

Thirdly, adverbs of likelihood such as 'possibly', 'probably', and 'definitely' as well as of magnitude like 'slightly', 'somewhat', 'rather', 'pretty', 'quite', 'decidedly', 'unusually', 'very', and 'extremely' are shown to be acquired progressively - Hoffner, Cantor, and Badzinski (1990) researching likelihood adverbs and Bashaw and Anderson (1968) magnitude adverbs.

The fourth category of abstract nouns include such words as 'benefit', 'challenge', 'courage', 'decision', 'enjoyment', 'freedom', 'kindness', 'longevity', 'opinion', 'pleasure', and 'respect.' Acquisition of such nouns was explored by Nippold, Ward-Lonergan, and Fanning (2005a) by calculating how often children use abstract nouns in their persuasive writing, with the result that such lexical items were shown to be used progressively more as children moved up in grade.
The fifth type of vocabulary item is literate verbs which consist of metalinguistic verbs (e.g., 'assert', 'concede', 'imply', 'predict', 'interpret', and 'confirm') and metacognitive verbs (e.g., 'remember', 'doubt', 'infer', 'hypothesize', 'conclude', and 'assume'). Astington and Olson (1987) found that comprehension of literate verbs gradually improved as children got older. The mean accuracy went from 45%, 42%, 59%, 71%, and 92% in Grades 6, 8, 10, 12 (N=99), and college students (N=77), respectively. Walton (2000) looked into spontaneous discourse of children in kindergarten through Grade 4 in terms of epistemological expressions such as verbs of certainty (e.g., 'know', 'believe', 'think', 'guess', 'expect', and 'sure'), verbs of veracity (e.g., 'to tell the truth', 'kid', 'fool', 'lie', 'trick', 'pretend', and 'tease'), and verbs of cognitive process (e.g., 'remember', 'forget', and 'figure out'). The children's use of epistemological expressions proved to increase in relationship to age. Spontaneous use of metacognitive and metalinguistic verbs in formal written language was examined by Nippold et al. (2005a) through verbs of 'argue', 'assert', 'disagree', 'discover', 'intend', and 'reflect' in 180 children, adolescents, and adults (ages 11, 17, and 24 years). An age-related increase was found in that 1.29 such words were used per 100 words in 11-year-old children while 1.90 in 17-year-olds and 1.95 in 24-year-olds were observed.

Lastly, factive (e.g., 'know', 'see', and 'notice') and nonfactive verbs (e.g., 'think', 'guess', and 'believe') have been explored in developmental studies. Scoville and Gordon (1980), for instance, examined comprehension of the factive verbs/words 'know', 'forget', 'sorry', 'happy', 'surprised' as well as the nonfactive verbs/words 'think', 'sure', 'figure', 'say', and 'believe' in 76 children (aged 5, 8, and 11), adolescents (aged 14), and adults (aged 20 years). The results showed that the knowledge of factive and nonfactive verbs is obtained later on.

The findings of these studies seem to indicate that understanding the subtle meaning of literate words takes a long time, even into one's early adulthood to acquire, though small children may use many of these words in
their spontaneous speech. Thus, as McNeil (1970) argues, simple counting of how many words a child can use may give a misleading picture of lexical development.

Turning our focus from the acquisition of literate words *per se* to lexical restructuring and retrieval in the brain, the literature (e.g., Nagy and Herman, 1987; Watson, 1985) shows that as new words are added to the lexicon, old words take on new and subtle meanings, and it becomes easier for people to organize and reflect upon the content of language held in the lexicon. Some neurolinguists such as Paradis (2004) and Bjork and Bjork (1992) argue (1) that the human capacity for long-term storage is virtually unlimited but retrieval capacity is limited and highly dependent on how frequently an item is retrieved and how recently learning took place, and (2) that "the act of retrieving an item of information is considerably more powerful in terms of facilitating its subsequent successful recall than is an additional study trial on that item" (Bjork and Bjork, 1992: 37). Lexical retrieval problems reveal themselves in the dysfluency phenomena of pauses, hesitations, circumlocutions, and the use of indefinite pronouns (e.g., something), empty fillers (e.g., thing, stuff), or lexical substitutions (e.g., animal for deer) during spoken communication in language disorder studies (e.g., Dennis, 1992; Snyder and Downery, 1991; Wiig, Zureich, and Chan, 2000) as well as in studies with normal subjects (e.g., Bjork and Bjork, 1992; Burke, MacKay, Worthley, and Wada, 1991). Such developmental studies also revealed (1) that accuracy and speed of naming gradually increase during the school-age and adolescent years, and well into adulthood and (2) that skill in word finding is closely associated with oral and silent reading abilities (Nippold, 2006).

Apart from the word list we include for lexical development analysis, attention is also paid to a conventional tool to detect lexical density or diversity, that is, TTR (type token ratio) which is calculated by the ratio of different words (type) to total words (token) in a text. In applying TTR to
research, however, caution has been suggested regarding text length. Reed (2000: 201) states on calculating lexical density "the figure obtained for some of the statistics varies according to the length of the text." Hess et al (1986) reported on spoken data from elementary school children that a text length of 50 to 100 words is not of sufficient length to obtain a reliable type-token ratio. Similar results are drawn from Richard et al.'s (1987) findings where a text consisting of less than 200 words does not yield reliable statistics in terms of lexical diversity. Thus, our study includes only those texts which contain more than 200 words, whether they are elicited as spoken or written data.

In the application of TTR to bilingual studies, Kimura (1994) calculated an overall ratio as a whole whereas Yoshitomi (1999) calculated TTR for open class words vs closed class words (open class vocabulary includes nouns, verbs, adjectives, and adverbs; closed class vocabulary includes articles, pronouns, and conjunctions), yielding the finding that the size of open class productive vocabulary shrinks as a function of incubation. Considering this result as well as Bierling's (1990) and Cohen's (1989) results that open class words are more attrition prone, calculating TTR for the two separate categories seems more appropriate for observing lexical changes over time. Examining attrition in the productive lexicon, Cohen (1989) further calculated TTR by word class (nouns, verbs, adjectives, adverbs, prepositions, conjunctions, pronouns, and interjections) and found a susceptibility in nouns.

The use of TTR is not without criticism as a tool for measuring lexical diversity, however. The TTR has been criticized as being affected by the number of words in the sample - a sample containing a large number of tokens gives a lower TTR (McKee, Malvern, and Richards, 2000). In order to adjust the sample size for a valid and reliable measure of vocabulary diversity, "CLAN" has a mathematical function called vocd, which was also used in this study. "CLAN" is one of the functions of CHILDES (Child
Bilingual First Language Development of Two Japanese-English Bilingual Siblings: Literature Review on Possible Dependent Variables

Language Data Exchange System) by McWhinney (1995), and widely used in bilingual studies. This software automates various measurements from transcripts to the Type-Token Ratio (TTR) and \textit{vocd}. Thus, in calculating TTR for various categories of words along with the total TTR, we also include \textit{vocd}.

Another tool worth noticing is the Lexical Frequency Profile (LFP) proposed by Laufer and Nation (1995: 311) which "shows the percentage of words a learner uses at different vocabulary frequency levels." Laufer and Nation use four levels of word lists: the first 1,000 most frequent words, the second 1,000 most frequent words, academic vocabulary, and words not in any of the above three lists. The computer program package \textit{Frequency Level Checker} (http://language.tiu.ac.jp/flc/) calculates which level each word belongs to as data texts are fed into the program. The results may reveal lexical changes and such lists have been previously used in bilingual studies (e.g. Keshavarz and Astaneh, 2004).

Although evidence of the validity and reliability of lexical frequency as a measure of vocabulary size was presented by Laufer and Nation (1995), Laufer (1994), in examining vocabulary development in L2 writing, reported that a clearer picture of an increase in productive vocabulary could be provided by collapsing Levels 1 and 2 and Levels 3 and 4. In line with this recommendation we included collapsed counts of Levels 1 and 2, and Levels 3 and 4 in both type and token. Thus, in our study, oral and written data with more than 200 words will be quantitatively analyzed to obtain the TTR. The analysis includes (1) an overall ratio, (2) an open vs closed class words ratio and (3) a ratio of nouns against the total tokens. Additional quantification involves the LEP calculation on the \textit{Frequency Level Checker} program.

3.2. Metaphors, similes, and idioms

Special attention is paid to certain lexical item in this study, that is, metaphors, similes, and idioms. Development of metaphors and similes (A is
like B) has been investigated (e.g., Gardner et al. 1975; Kogan et al., 1980; Reynolds and Orthony, 1980) and it has been found that understanding of metaphors and similes continues to deepen during the school-age years and adolescence and into adulthood in both monolingual and bilingual children (Nippold, 2006).

Qualls and Harris (1999) examined idiom understanding in a group of 48 children (10-year-old African American children and 10-year-old European American children) living in the mid-southern region of the USA, and found that the two cultural groups performed similarly on some of the expressions but differed on others. Developmental studies have shown that understanding of idioms gradually improves during the school-age and adolescent years and even into adulthood (Nippold, 2006). Yorio (1994: 68), in examining conversationalized language (idioms, formulas, prefabricated patterns) by adult second language learners, found that acquiring such language forms is "a long and arduous task." We investigate whether bilingual children's acquisition of metaphors, similes, and idioms resembles that of monolinguals or adult second language learners. In the investigation we use Nippold's list of 100 idioms (2006: 204-207), one example of which is 'beat around the bush'.

3.3 Morphology

One type of English morphology is called 'inflectional morphology' which simply adds, for instance, suffixes such as the plural marker 's' or the past tense marker 'ed' in extending the meanings of the original word. This morphology is generally mastered by the age of six or seven years whereas the second type, derivational morphology, continues to develop throughout school-age and well into adulthood (Berko, 1985; Derwing and Baker, 1979; Levin, Ravid, and Rapaport, 2001). Examples of derivational morphology are the noun 'happiness' derived from the adjective 'happy' or the noun 'reception' derived from its verb 'receive' ('reception' is 'non-neutral' in that it
undergoes a phonetic change).

Developmental studies on children's L1 English derivational morphology have been extensively carried out with the following four major findings: (1) important developmental changes take place between kindergarten and Grade 1 due to children's exposure to literacy education (Levin et al., 2001), (2) receptive tasks are easier than productive tasks and non-neutral derived forms are more difficult to acquire than neutral forms (Carlisle and Nomanbhoy, 1993), (3) an explosive growth in derivational development occurs between ages 9 and 14 with increased exposure to written language which contains many morphologically complicated words compared to the spoken language (White et al., 1989), and (4) knowledge of derivational morphemes is strongly associated with spelling - the two skills improve with each successive grade level (Carlisle, 1988: Green et al, 2003). These findings of L1 developmental studies on derivational morphology appear to indicate that reading, and spelling is inseparable and has to be cautiously treated as a whole, when looking at the children's use of derivational morphology.

When examining the actual derivational morphological forms to test children's developmental acquisition, (1) noun suffixes such as 'ist', 'ity', 'ion', and 'ation', (2) verb suffixes like 'ify', 'ize', and 'ate', and (3) adjectives ending in 'ous', 'ious', 'ive', and 'al' were used by Mahony (1994) while Mahony et al. (2000) used such derivational pairs as 'person-personal', 'space-spatial', and 'nature-natural' in contrast to the non-derivational pairs of 'cat-catalogue' and 'general-generous'. These suffixes, adjectives and pairs are used in investigating the developmental acquisition of derivational morphemes in this study.

Our attention is now turned to a theoretical construct of the study for morphological analysis, and the inclusion of the 4-M (morpheme) model (Figure 1) and its analysis is considered. Based on Levelt's (1989) speech production model, Myers-Scotton (2002) puts forward the 4-M model which
primarily attempts to explain how morphemes are elected (activated) in the speech production process. The 4-M model proposes four different types of morphemes (content, early system and two late system morphemes) against the traditional free-bound morpheme distinction. Different lemmas underlying each type of morpheme become salient at different stages of speech production.

**Figure 1.** Myers-Scotton's (2002) production process diagram: lemma activation

Content morphemes, which are directly elected by a speaker's intentions to map conceptual structure onto the lemma, assign or receive a thematic role. Early system morphemes pattern with content morphemes through indirect election by content morphemes, but they do not assign or receive a thematic
role. While content and early system morphemes are conceptually activated, the two other types of system morphemes are structurally assigned "later" in the production process when lemmas send directions to the Formulator. These morphemes are called bridge and outsider late system morphemes. Examples of early, bridge and outsider system morphemes are the definite article 'the', possessive preposition 'of', and third person singular $s$ respectively. The definite article 'the' is conceptually activated to connect the content of the semantic and pragmatic feature bundle, definiteness in this case, to its head noun. The bridge late morphemes are activated when grammatically correct forms call for them, not depending on the head of the maximal projection but on other information in this maximal projection (a maximal projection is the highest projection of a lexical head such as VP for $V$, and NP for $N$). Meanwhile, the outsider late system morphemes "look outside the maximal projection in which they occur" (Myers-Scotton, 2002: 18). The subject-verb agreement such as a third person singular $s$ is therefore classified as such.

Myers-Scotton's dynamic model is relevant to our study since it can be used to explain or predict language acquisition stages. The model assumes "that learning a language depends on mapping conceptualizations onto an abstract lexicon and the grammar it projects" (Myers-Scotton and Jake, 2000: 1087) and hypothesizes that (1) conceptually activated content morphemes are acquired before system morphemes, (2) early system morphemes are acquired before late system morphemes, and (3) bridge morphemes are produced more accurately than outsiders. Wei (2000) provided empirical support for these predictions when he examined adult Chinese/Japanese ESL learners ($N=60$) in the USA in light of developmental stages of English morpheme acquisition. Furthermore, using Wei's data, Myers-Scotton (2002: 1088) successfully showed inadequate use of free-bound morpheme distinction and the validity of her model (classifying morphemes into four types) by paying special attention to the three types of the English morpheme
s - plural s, possessive 's, and 3rd person present tense s. Wei's data support the model in that the early system morpheme (plural s) is more accurately produced than the structurally activated late system morphemes. Within the category of late morphemes, the bridge morpheme (possessive 's) induces less errors than the outsider morpheme (subject-verb agreement s).

Thus, the 4-M model seems capable of serving as a useful tool to detect developmental stages of morphological acquisition by distinguishing four different morpheme types rather than conventional content vs. function dichotomy. As for the specific morphemes we examine, 36 variables are taken from Taura's (2008) study on Japanese-English bilinguals (Table 8).

3.4 Syntax

Developmental studies on syntactic maturity show that frequent opportunities to read and write in literate ways result in syntactic growth. As Perera (1992) claims, children are exposed to low frequency syntactic structures (which they may not otherwise encounter in casual oral communication settings) through literal reading and they are then able to practice such structures through formal writing at school. Syntactic maturity takes a long time, although by early adulthood a great deal of syntactic sophistication has been achieved through constant reading and writing at school. Even well-educated adults find themselves not having mastered certain aspects of syntax (Menyuk, 1977), which is exemplified in their speech as maze (e.g., false starts, hesitation, or revisions). Therefore, in researching children's syntactic development, it is desirable for research focus to be placed both on accuracy and temporal aspects of speech. Meanwhile, dialectical differences in school-age children are shown to exert little effect on syntactic development which is gradual and subtle (Smith, Lee, and McDade, 2001). This finding is useful to us in that we can compare the literature (mostly concerned with American or British dialects of English) with data taken from the bilingual siblings whose mother is Australian.
Nippold (2006), in her overview of developmental studies on syntax in monolingual children, categorizes the literature into two types, depending on whether a study is concerned with intrasentential level (within sentence boundaries) or intersentential level (between sentence boundaries). First, research that examines intrasentential level focuses (1) on sentence length: Loban (1976) found the number of words (or subordinate clause) in each C-unit both in spoken and written data becomes lengthy as a function of grade (ages 6 through 18) as shown in Table 1, while Cuterrez-Clellen and Hofstetter (1994) demonstrated the use of subordinate clauses as an important part of attaining syntax by examining the syntactic complexity of narratives (Frog Goes to Dinner) as told by 77 Spanish-speaking children ages 5, 6, and 8 years as seen in Table 2, (2) on discourse genre which is claimed to be a factor that greatly influences syntactic complexity: Leadholm and Miller (1992) report in Table 3 that children produce longer MLUs in a narrative than in conversation while Crowhurst (1980) finds T-unit length significantly longer in persuasive writing than in narrative writing (Table 4), (3) low frequency syntactic structures: uncommon structures (e.g., appositives, elaborated subjects, postmodification via prepositional phrases, and non finite verbs) and complex and lengthy verb phrases (e.g., modal auxiliary verbs, the perfect aspect, and the passive voice) are progressively used as children grow older (Scott, 1988b, and Scott and Stokes, 1995), (4) the order of two clauses: sentences comprised of a main clause followed by a subordinate (adverbial) clause are much more frequently used by children than those of a subordinate clause preceding a main clause, implying left-branching structures are more difficult to acquire (Loban, 1976; Scott and Stokes, 1995), and (5) the type of cohesion device - whether conjunctions are subordinating such as 'when' and 'although', coordinating like 'and' and 'but', or correlative like 'both' and 'either': the literature indicates that the former two types of conjunctions increase in frequency of use as a function of grade level (Flores d'Arcais, 1978; Savage and Fallis,
1988; Wing and Scholnick, 1981) but not much research has been conducted on correlative conjunctions to date. In addition, we include the number of relative clauses per C-/T-Unit since one of our subjects' first languages is Japanese which does not have an equivalent form, and is found to be difficult to master by Japanese learners of English (e.g. Flynn, 1994).

### Table 1

<table>
<thead>
<tr>
<th>age</th>
<th>spoken</th>
<th>written</th>
<th># words/C-Unit</th>
<th># subordinate clause/C-Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-7</td>
<td>6088</td>
<td></td>
<td>0.16</td>
<td>1.18</td>
</tr>
<tr>
<td>8-9</td>
<td>7.62</td>
<td>7.6</td>
<td>0.22</td>
<td>0.29</td>
</tr>
<tr>
<td>11-12</td>
<td>9.82</td>
<td>9.04</td>
<td>0.37</td>
<td>0.47</td>
</tr>
<tr>
<td>14-15</td>
<td>10.98</td>
<td>10.05</td>
<td>0.43</td>
<td>0.47</td>
</tr>
<tr>
<td>17-18</td>
<td>11.7</td>
<td>13.27</td>
<td>0.58</td>
<td>0.6</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>age</th>
<th># words</th>
<th># clause</th>
<th># relative</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6.5</td>
<td>1.11</td>
<td>0.39</td>
</tr>
<tr>
<td>6</td>
<td>6.9</td>
<td>1.18</td>
<td>1.1</td>
</tr>
<tr>
<td>8</td>
<td>7.3</td>
<td>1.23</td>
<td>1.6</td>
</tr>
</tbody>
</table>

### Table 3. MUL in morphemes by Leadholm & Miller (1992)

<table>
<thead>
<tr>
<th>age</th>
<th>conversation</th>
<th>narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>5:04</td>
<td>5.71</td>
<td>&lt; 6.06</td>
</tr>
<tr>
<td>7:01</td>
<td>5.92</td>
<td>&lt; 7.32</td>
</tr>
<tr>
<td>9:01</td>
<td>6.5</td>
<td>&lt; 8.8</td>
</tr>
<tr>
<td>11:01</td>
<td>7.62</td>
<td>&lt; 9.83</td>
</tr>
<tr>
<td>13:00</td>
<td>6.99</td>
<td>&lt; 9.32</td>
</tr>
</tbody>
</table>

### Table 4. Crowhurst & Piche's (1979)

<table>
<thead>
<tr>
<th>age</th>
<th>conversation</th>
<th>narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>5:04</td>
<td>5.71</td>
<td>&lt; 6.06</td>
</tr>
<tr>
<td>7:01</td>
<td>5.92</td>
<td>&lt; 7.32</td>
</tr>
<tr>
<td>9:01</td>
<td>6.5</td>
<td>&lt; 8.8</td>
</tr>
<tr>
<td>11:01</td>
<td>7.62</td>
<td>&lt; 9.83</td>
</tr>
<tr>
<td>13:00</td>
<td>6.99</td>
<td>&lt; 9.32</td>
</tr>
</tbody>
</table>

Research on intersentential level, on the other hand, mainly concentrates on adverbial conjunction (e.g., 'accordingly' and 'thus'). Testing accurate use of ten conjunctions ('consequently', 'contrastively', 'conversely', 'furthermore', 'however', 'moreover', 'nevertheless', 'rather', 'similarly', 'therefore') on 120 adolescents and young adults (30 in each of the age groups of 12, 15, 19, and 23 years), Nippold et al. (1992) found that performance on both tasks steadily improved with increasing age. Steffani and Nippold (1997) examined the use (writing) and understanding (reading) of the same set of adverbial conjuncts in 20 non-native speakers of English (mean age of 23
years) whose native language was Japanese, and found that the accuracy rate correlated with their length of stay in the USA. Spontaneous conversation data naturalistically collected from 60 children, adolescents, and adults (20 in each age group of 11, 17, and 25 years) were examined by Nippold et al. (2005b) with the finding that adults used such conjunction words and phrases more frequently as 'actually', 'basically', 'definitely', 'essentially', 'eventually', 'for example', 'generally', 'literally', 'normally', 'obviously', 'originally', 'otherwise', 'personally', 'typically', 'unfortunately', and 'usually'.

3.5 Speaking (conversation and narration)

Oral data have been collected from a number of bilingual studies to examine productive skills based on the assumption that it requires on-line processing under time pressure. Some researchers have used pictures to elicit data (Berman & Olshtain, 1983; Cohen, 1989; Hansen-Strain, 1990; Kuhberg, 1992; Olshtain, 1989; Tomiyama, 2000, 2009) while others have collected spontaneous speech samples through free conversation in natural settings (Kravin, 1992; Littlewood, 1996; Nakuma, 1997; Yoshitomi, 1999). We'll focus on picture elicitation data first, followed by conversation data.
### Table 5. Temporal variables

<table>
<thead>
<tr>
<th>Temporal variables</th>
<th>Pause distribution variables</th>
</tr>
</thead>
<tbody>
<tr>
<td># Items</td>
<td># Items (pause&lt;200 ms.)</td>
</tr>
<tr>
<td>total duration (sec)</td>
<td>% # intra T-unit pause / total # pause</td>
</tr>
<tr>
<td>total speaking duration (sec)</td>
<td>% # inter T-unit pause / total # pause</td>
</tr>
<tr>
<td>total pause duration (sec)</td>
<td>% inter T-unit pause / total pause duration</td>
</tr>
<tr>
<td>total # pauses + # zero-paused inter T-unit boundaries</td>
<td>% intra T-unit pause / total pause duration</td>
</tr>
<tr>
<td># inter T-unit boundaries</td>
<td>% inter T-unit pause / total duration</td>
</tr>
<tr>
<td>total # T-units</td>
<td>% intra T-unit pause / total duration</td>
</tr>
<tr>
<td># inter T-unit zero-paused boundaries</td>
<td>% # inter T-unit pause / # T-unit boundaries</td>
</tr>
<tr>
<td># intra T-unit pause</td>
<td></td>
</tr>
<tr>
<td># inter T-unit pause (excl. zero-paused)</td>
<td></td>
</tr>
<tr>
<td>total # pause longer than 200 ms</td>
<td></td>
</tr>
<tr>
<td>average inter T-unit pause (ms)</td>
<td></td>
</tr>
<tr>
<td>average inter T-unit pause (excl. zero)</td>
<td></td>
</tr>
<tr>
<td>total intra T-unit pause duration</td>
<td></td>
</tr>
<tr>
<td>average intra T-unit pause (ms)</td>
<td></td>
</tr>
<tr>
<td>% speaking / total duration</td>
<td></td>
</tr>
<tr>
<td>% total pause duration / total duration</td>
<td></td>
</tr>
<tr>
<td>average # words / minute</td>
<td></td>
</tr>
<tr>
<td>average # T-units / minute</td>
<td></td>
</tr>
<tr>
<td>average # clauses / minute</td>
<td></td>
</tr>
<tr>
<td>average # morphemes / minute</td>
<td></td>
</tr>
<tr>
<td>average # morphemes / minute (speaking duration)</td>
<td></td>
</tr>
<tr>
<td>average # EF T-units / minute</td>
<td></td>
</tr>
<tr>
<td>average # EF T-units / minute (speaking duration)</td>
<td></td>
</tr>
<tr>
<td>average # fillers / minute</td>
<td></td>
</tr>
<tr>
<td>average # false start / minute</td>
<td></td>
</tr>
<tr>
<td>average # words in false start / minute</td>
<td></td>
</tr>
<tr>
<td>average # abandon / minute</td>
<td></td>
</tr>
<tr>
<td>average # of wrong words / minute</td>
<td></td>
</tr>
<tr>
<td>average # fillers / minute (speaking duration)</td>
<td></td>
</tr>
<tr>
<td>average # false start / minute (speaking duration)</td>
<td></td>
</tr>
<tr>
<td>average # words in false start / minute (speaking duration)</td>
<td></td>
</tr>
<tr>
<td>average # abandon / minute (speaking duration)</td>
<td></td>
</tr>
<tr>
<td>average # of wrong words / minute (speaking duration)</td>
<td></td>
</tr>
</tbody>
</table>
Among the pictures and picture books used by researchers in the first group of studies, a series of ‘frog’ storybooks by Mayer (1969) has often been utilized as a way of eliciting narratives that encourages narrators to undertake the task of putting simultaneous activities into order (Berman and Olshtain, 1983; Cohen, 1989; Olshtain, 1989; Olshtain and Barzilay, 1991; Taura, 2005 and 2008; Yoshitomi, 1999). In relation to the 'Frog' series' age-appropriateness in bilingual studies, the ages vary from children of five to fourteen (Berman and Olshtain, 1983; Cohen, 1989; Olshtain, 1989; Yoshitomi, 1999) to adults (Olshtain and Barzilay, 1991; Yagmur, 1997). Thus it seems that the data obtained using the 'Frog' series are readily comparable to that reported in the existing literature for lexical and morphosyntactic analysis. The data are to be analyzed from lexical, accuracy, complexity, and fluency viewpoints. Through reviewing the literature (i.e., Scott and Windsor, 2000; Taura, 2005 & 2008; Wolfe-Quintero, Inagaki, and Kim, 1998), for complexity measurements, we calculate the percentage of the number of complex clauses per clause (complex clauses include embedded relative clauses and subordinate clauses as well as clauses with passivity). Fluency will be examined in this study by counting the total number of words, clauses, and T-units along with 44 pause-related variables as listed in Table 5.

Brinton and Fujiki (1984) examined the development of conversation (spontaneous dinner table) data taken from 36 individuals - 2 groups of school-age children (ages 5 and 9 years) along with one group of young adults. The results (Table 6) showed (1) the adults' significantly longer maintenance of the same topic and number of utterances per topic, along with more topic shading (development of the ability to stay on topic and shift gracefully from one topic to another) and (2) large individual differences within age groups - differences in interpersonal style and personality may affect the types of conversational behaviors. Larson and McKinley's (1998) study also disclosed individual or gender differences that boys made
significantly more attempts than girls to entertain their interlocutors using jokes (3.47 and 1.33, respectively).


<table>
<thead>
<tr>
<th></th>
<th>Mean # of introduction</th>
<th>Mean % reintroduction</th>
<th>Mean # of maintenance</th>
<th>Mean # of utterance/topic</th>
<th>Mean # of topic shading</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>23.50</td>
<td>22.83</td>
<td>79</td>
<td>5.08</td>
<td>4.83</td>
</tr>
<tr>
<td>9-year-olds</td>
<td>23.33</td>
<td>20.50</td>
<td>84</td>
<td>6.34</td>
<td>6.00</td>
</tr>
<tr>
<td>adults</td>
<td>13.17</td>
<td>6.00</td>
<td>96</td>
<td>10.69</td>
<td>10.50</td>
</tr>
</tbody>
</table>

Turning to studies on narration, next, evidence suggests that narrative ability is associated with reading achievement (e.g., Klecan-Aker and Caraway, 1997). Narration is thought to be a more challenging and sophisticated discourse genre than conversation. As MacLachlan and Chapman (1988: 2-3) explained, narration demands a more complex syntax, offers less discourse support, and poses more organizational problems than conversation.

Research shows that as children mature, the following development occurs: (1) stories gradually become longer, more detailed, and better organized, (2) stories contain a greater number of episodes, (3) episodes are more likely to be completed and embedded within larger episodes, that is, sub plotting, (4) cohesion across episodes increases through the use of syntactic devices such as conjunctions, (5) more is said about the characters' emotions, thoughts, and plans, and (6) greater effort is made to entertain and engage the listener (Bamberg and Damrad-Frye, 1991; Botvin and Sutton-Smith, 1977; Crais and Lorch, 1994; Liles, 1993; Strong and Shaver, 1991).

Inner thoughts of the characters were investigated by Bamberg and Damrad-Frye (1991) by using 'Frog, where are you?' to 36 participants (children ages 5 and 9 years along with 20-year-old adults). The findings were (1) the stories told by the adults were longer than those narrated by the children (the 20-year-olds produced an average of 79.25 clauses per story
while it was 47.58 with the 5-year-olds and 46.08 clauses with the 9-year-olds), (2) when differences in story length were controlled, the adults were more likely to mention the thoughts, feelings, and emotions of the characters through the use of words such as 'happy', 'sad', and 'scared' (of the total number of clauses produced by the adults, 20% referred to mental states while only 4% and 7% for the 5-year-olds and 9-year-olds, respectively), reflecting adults' greater awareness of other people's minds which is consistent with Dorval and Eckerman's (1984) study. It has been reported by many studies (e.g., Bamberg and Dmarad-Frye, 1991; Kernan, 1977; Stein and Glenn, 1979) that as children grow older, they are more likely to mention the thoughts and emotions of story characters.

Liles (1987) examined other aspects of narrative ability and found an age-related improvement (the use of conjunctions across episodes to create cohesive discourse increased in relation to age) when 20 children aged 7 to 10 were studied in terms of their ability to use appropriate conjunctions expressing additive ('and', 'also'), temporal ('meanwhile', 'next'), causal ('so', 'therefore'), and adversative ('yet', 'but') relations, which was consistent with the view of Mentis (1994) and Scott (1984) that students gradually acquire skills in using adverbial conjunctions such as 'anyway', 'meanwhile', 'incidentally' and related phrases such as 'before I forget' or 'that reminds me'.

3.6 Writing

Development of reading and writing is a gradual process and the key to developing these skills is exposure to literate texts where such advanced syntactic structures are employed as subordinate clauses, multiple embedding, the passive voice, and other low-frequency devices. Unlike the oral skills of speaking or listening, children need to receive systematic instruction in reading and writing to gain high levels of proficiency because adults can serve as "scaffolds" to coin the Vygotskian term (1986).

Perera (1986) argues that "as children and adolescents gain knowledge of
complex syntax, literate vocabulary, and figurative expressions that they encounter at school when reading books or listening to their teachers' lectures, they incorporate these features into their own essays as they think about the topic and decide how best to express their own thoughts, observations, and insights." Not much research has been carried out on this issue of children's exposure to literate texts.

As two of the main components of writing skills (other than morphology and syntax which were examined earlier), research on spelling and learning to write is to be overviewed here.

First, as Ehri (2000) claims, reading and spelling are like two sides of a coin in that both rely on the same knowledge sources in the memory: knowledge about the alphabetic system and knowledge about the spelling of specific words. Once again, however, like reading, spelling requires explicit instruction, frequent opportunities to practice what has been learned, and feedback from teachers and parents (Ehri, 2000; Moats, 1995).

The second component is learning how to write: Children in English speaking countries undergo certain stages in this process (1) firstly at kindergarten reciting the alphabet (Open Court Reading, 2000), (2) reviewing the sounds, beginning to write independently, and learning how to spell common words in Grade 1 (ages 6-7 years), (3) increasing their writing volume dramatically by Grade 3 (ages 8-9 years) by combining simple and complex sentences into paragraphs for an organized and clear composition style.

Both spelling and writing need guidance from adults, and a query arises here as to whether the feedback provided to bilingual children is insufficient or not and this is explored in this study as well.

Next, we briefly review the L2 writing research with the view to finding an appropriate tool for our research. L2 writing research to date has covered many aspects of L2 compositions. Jacobes et al (1981) researched content, organization, vocabulary, language and mechanics. Mullen (1977) looked at
control over English structure, organization of material, appropriateness of vocabulary, quantity of writing, and overall writing proficiency (a composite rating derived by adding the scores of the first 4 categories). Pritchard (cited in Brown and Bailey, 1984: 26) studied organization, length, content, grammar, vocabulary, and spelling while Hedgcock and Lefkowitz (1992) considered grammar, vocabulary, and mechanics.

Although most studies here have dealt with both ‘product (linguistic accuracy)’ and ‘process (organization) ’ of L2 (ESL or EFL) writing, Polio (1997: 103) points out that in current writing pedagogy there is an overemphasis on the writing process or idea generation rather than simply getting students to write error-free sentences. In the same vein, Silva (1993) calls for further research to fill in the gaps of previous studies with a more even distribution of strategic, rhetorical and linguistic concerns. Similarly, Uzawa (1996), who examined the characteristics of skilled and unskilled writers, contends that unskilled writers are overtly concerned with mechanics such as spelling, punctuation, and grammar. She further argues that unskilled writers tend to cut short the planning phase before writing and neglect organization and the audience. The present study therefore seeks to explore both the process and product of L2 writing by examining the organization and content as well as grammar and mechanics, including the punctuation, spelling and vocabulary of writing samples.

We turn our attention to scoring measurements of collected data. The question of global ratings versus more objective scoring methods has been debated in the field of L2 writing. Brown and Bailey (1984) modified Stiggins’ (1982) approach to scoring L2 writing by adding one more recent quantifying approach, that of holistic scoring in which “each composition is assigned a unitary rating, often on a four to eight-point scale” (p22). An example of this type of holistic scoring can be seen in Hamp-Lyons and Henning (1991), where they measure linguistic accuracy on a nine-point scale: The maximum of nine is given when the reader sees no errors in
vocabulary, spelling, punctuation, or grammar while a minimum of five is given when errors frequently intrude.

Analytic scoring is a second type of global rating which “involves isolating one or more characteristics of writing and scoring them individually. Such characteristics may include style, grammar, mechanics, etc.” (Brown and Bailey, 1984: 22). A third rating guideline employs objective (quantifying) scoring by using such tools as “T-units or a ratio of cohesive devices to composition length” (Brown and Bailey, 1984: 22). The notion of T-unit (terminable unit; the smallest unit within a sentence that can stand alone), which was first introduced by Hunt (1965), has been employed in many other L2 writing studies such as Flahive and Snow (1980), Kameen (1979), and Perkins (1980). Hunt (1965) found that the ratio of clauses per T-unit and the length of the T-unit increases as writing skills develop. Flahive and Snow (1980) found strong correlations between holistic scores and T-unit measures with higher level ESL learners.

In an empirical study comparing major measures used in L2 writing research, Polio (1997) categorized the most frequently used measurements into three methods: (1) holistic (analytic scoring is included here), (2) error-free T-units, and (3) an error classification system. In their quest to find out which method, either direct (holistic and analytical) or indirect (multiple choice and objective tests) was more reliable, Perkins, Pohlmann, and Brutten (1988) found the holistic/analytical method to be more reliable than objective assessment.

We considered a number of these different scoring measures used in previous studies in order to find more reliable measures to use in this study. It was decided to adopt the holistic/analytical measures for two main reasons: first, because of the findings by Perkins et al (1988) mentioned above, and second, because of Ishikawa’s (1995) research showing that holistic/analytical measures are aimed at placement, and are more suitable when dealing with a wide range of proficiencies than when evaluating
homogeneous populations at a single proficiency level. Both a
sub-categorical item analysis (analytical) and overall assessment (holistic)
are required when looking at multi-dimensional aspects of writing.

Based on the discussion above, we looked for a tool which best met Perkin
et al’s (1988) criteria (as well as taking into account the time pressure factor)
and decided to use the third Edition of the Test of Written Language
(TOWL-3) created by Hammill and Larsen (1996). Hammill and Larsen
claim that TOWL-3 looks at three abilities in writing. First, it examines the
ability to write in compliance with accepted standards, especially those
governing punctuation, capitalization, and spelling. These skills are tested
within the subset known as Contextual Conventions (CC). Secondly, the
Contextual Language (CL) rating describes the subject’s ability to use the
syntactic, morphological, and semantic elements of English (suitable
wording, tense, plural, noun-verb agreement). Lastly, the Story Construction
subset (StC) analyzes the subject’s ability to express ideas, opinions and
thoughts in a creative and mature way (i.e. logical and coherent story
development and reader impact). Thus, TOWL-3 seems to satisfactorily
cover all the aspects of L2 writing this study aims to examine.

Moreover, TOWL-3 includes other relevant factors such as introductory
phrases, interrogative sentences, coordinate and subordinate coordination,
and articles. Silva (1993) mentions the significance of introductory phrases
which tend to be few in number and small in range in L2 learners’ writing
samples. Dunkelblau (1990) notes that interrogative sentences appear less
frequently in L2 writing. Silva (1993) and Elliot (1986) found more
subordinate coordination and Oi (1984) discovered fewer articles in L2
writing.

With reference to scoring measures, this study adopts holistic/analytic
methods (with the total word count being the only objective measure) as
mentioned above. TOWL-3 uses a two to-four-point scoring scale for all
items across the three subsets CC, CL and StC. Each subset employs an
analytical approach for discrete items such as “subject-verb agreements in the CL subset (0 points for more than one error, 1 point for one error, or 2 points for no errors), as well as a holistic approach, evaluating prose in the StC subset, for example, as immature (0 points), ordinary (1 point), or stylish (2 points). There are 11-14 items in each subset; the CC, CL, and StC raw scores are individually added up and then transformed into the standard scores provided by TOWL-3 for the respective subsets (See Table 7).

Table 7. Guidelines for Interpreting Standard Scores in CC, CL, and StC

<table>
<thead>
<tr>
<th>Standard Score</th>
<th>Description</th>
<th>% Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 to 20</td>
<td>Very Superior</td>
<td>2.34</td>
</tr>
<tr>
<td>15 to 16</td>
<td>Superior</td>
<td>6.87</td>
</tr>
<tr>
<td>13 to 14</td>
<td>Above Average</td>
<td>16.12</td>
</tr>
<tr>
<td>8 to 12</td>
<td>Average</td>
<td>49.51</td>
</tr>
<tr>
<td>6 to 7</td>
<td>Below Average</td>
<td>16.12</td>
</tr>
<tr>
<td>4 to 5</td>
<td>Poor</td>
<td>6.87</td>
</tr>
<tr>
<td>1 to 3</td>
<td>Very Poor</td>
<td>2.34</td>
</tr>
</tbody>
</table>

* Full score=20, mean score=10, SD=3

For every one of the sub tests at each age level, the mean score is set at 10 with a standard deviation (SD) of 3. Normative data in TOWL-3 are provided based on 2,217 North American native English speaking children, aged 7-18, to give useful guidelines for grading each subject’s writing level. TOWL-3 also calculates the total sum derived by adding the three subset standard scores and converting the sum to a quotient using the table. This “Composite Quotient” estimates the subject’s overall writing competence, with a mean score of (M) 100 and standard deviation (SD) fixed at 15. Quotients of 131-165, for example, are described as “very superior” (top 2.34%), 90-110 as “average” (49.51%) and 80-89 “below average” (16.12%).

Both contrived and spontaneous formats are used in TOWL-3. The contrived format taps the discrete elements or the clearly defined boundaries of the language used and evaluates them in isolation. The spontaneous format on the other hand tests the ability to write freely about any topic
pertinent to the reader without boundaries. The studies mentioned earlier on (e.g. Polio, 1997; Wolfe-Quintero, Inagaki and Kim, 1998; Hedgcock and Lefkowitz, 1992; Silva, 1993; Uzawa, 1996) did not use the contrived format due to its unnatural way of examining the written language. This study therefore also adopts the spontaneous format alone, with the aim of employing a test format that is closer to a classroom situation where students are asked to write on a specific topic rather than being tested on a limited range of grammatical questions. Another reason for using only the TOWL-3 spontaneous format is that it requires examinees to write an expository essay. Silva (1993), in reviewing the existing literature, found an overall preference for researchers to use expository essays over argumentative and narrative types. Thus the spontaneous format allows an easier comparison of this study’s results with others’.

4. List of dependent variables

Through the literature review above, a large number of variables are found to have been used in L1 developmental studies. Our study has an explorative nature where bilingual siblings' language acquisition is examined in terms of whether or not the developmental stages are similar to those of monolingual English speakers, therefore an attempt is made to incorporate as many variables as possible. Below is the list of variables in each category.

4.1 Lexicon

The following words are examined in terms of developmental lexical acquisition in comparison with English monolingual children's acquisition.

4.1.1 Polysemous words: secondary meanings of 'above', 'low', 'up',
4.1.2 Double-function words: physical meanings of 'bright', 'cold', 'crooked', 'deep', 'hard', 'soft', 'sweet', 'warm'
4.1.3 Adverbs of likelihood and magnitude: 'decidedly', 'definitely',
'extremely', 'possibly', 'pretty', 'probably', 'quite', 'rather', 'slightly', 'somewhat', 'unusually', 'very'

4.1.4 Abstract words: 'benefit', 'challenge', 'courage', 'decision', 'enjoyment', 'freedom', 'kindness', 'longevity', 'opinion', 'pleasure', 'respect'

4.1.5 Literate words: 'argue', 'assert', 'assume', 'believe', 'concede', 'conclude', 'confirm', 'disagree', 'discover', 'doubt', 'expect', 'forget', 'figure out', 'fool', 'guess', 'hypothesize', 'imply', 'infer', 'interpret', 'intend', 'kid', 'know', 'lie', 'predict', 'pretend', 'reflect', 'remember', 'sure', 'tease', 'think', 'to tell the truth', 'trick'

4.1.6 Factive and nonfactive words: 'believe', 'figure', 'forget', 'guess', 'happy', 'know', 'notice', 'say', 'see', 'sorry', 'sure', 'surprised', 'think'

4.1.7 Dysfluency words: 'something', 'stuff', 'thing'

In addition to the words listed above, in examining the productive vocabulary from both the spoken and written data, we will take a quantitative approach by using the software CLAN and VocabProfile.

CLAN helps us to produce TTR related items: (1) the numbers of types and tokens of nouns, verbs, adjectives, adverbs, prepositions, conjunctions, and pronouns, (2) TTRs of nouns, verbs, adjectives, adverbs, prepositions, conjunctions, and pronouns, (3) a TTR comparison between open class words vs. closed class words, (4) TTR percentage of nouns, verbs, adjectives, adverbs, prepositions, conjunctions, and pronouns against the total TTR, and (5) the total TTR and vocd.

The Lexical Frequency Profile on the VocabProfile software provides us with the information on the particular level each word in the individual written and spoken samples belongs to. It also supplies the numbers of types and tokens in each level for individual samples, as well as the TTRs for each level. Further, we are able to calculate the percentage of types in the individual levels in terms of the total number of tokens contained in each
sample with the additional combination of percentages of Levels 1 and 2, and Levels 3 and 4. The procedure is repeated for tokens as well.

4.2 Metaphors, similes, and idioms

We use Nippold's (2006) list of 100 idioms to see if our bilingual subjects use idioms in the same developmental sequence as observed in English monolingual children.

4.3. Morphology

Table 8. 36 4-M model related variables

<table>
<thead>
<tr>
<th>#</th>
<th>morpheme types</th>
<th>examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>content morpheme</td>
<td>noun</td>
</tr>
<tr>
<td>2</td>
<td>content morpheme</td>
<td>pronoun</td>
</tr>
<tr>
<td>3</td>
<td>content morpheme</td>
<td>verb</td>
</tr>
<tr>
<td>4</td>
<td>content morpheme</td>
<td>auxiliary</td>
</tr>
<tr>
<td>5</td>
<td>content morpheme</td>
<td>adjective</td>
</tr>
<tr>
<td>6</td>
<td>content morpheme</td>
<td>adverb</td>
</tr>
<tr>
<td>7</td>
<td>content morpheme</td>
<td>free standing demonstrative (this is)</td>
</tr>
<tr>
<td>8</td>
<td>content morpheme</td>
<td>conjunction</td>
</tr>
<tr>
<td>9</td>
<td>content morpheme</td>
<td>relative</td>
</tr>
<tr>
<td>10</td>
<td>content morpheme</td>
<td>preposition</td>
</tr>
<tr>
<td>11</td>
<td>early system morpheme</td>
<td>article</td>
</tr>
<tr>
<td>12</td>
<td>early system morpheme</td>
<td>possessive pronoun</td>
</tr>
<tr>
<td>13</td>
<td>early system morpheme</td>
<td>demonstrative (this pen)</td>
</tr>
<tr>
<td>14</td>
<td>early system morpheme</td>
<td>plural 's'</td>
</tr>
<tr>
<td>15</td>
<td>early system morpheme</td>
<td>present participle</td>
</tr>
<tr>
<td>16</td>
<td>early system morpheme</td>
<td>past participle</td>
</tr>
<tr>
<td>17</td>
<td>early system morpheme</td>
<td>phrasal preposition</td>
</tr>
<tr>
<td>18</td>
<td>bridge late system morphemes</td>
<td>possessive 'of'</td>
</tr>
<tr>
<td>19</td>
<td>bridge late system morphemes</td>
<td>possessive 's'</td>
</tr>
<tr>
<td>20</td>
<td>bridge late system morphemes</td>
<td>formulaic article (all of a sudden)</td>
</tr>
<tr>
<td>21</td>
<td>bridge late system morphemes</td>
<td>dummy 'it'</td>
</tr>
<tr>
<td>22</td>
<td>bridge late system morphemes</td>
<td>dummy 'there'</td>
</tr>
<tr>
<td>23</td>
<td>bridge late system morphemes</td>
<td>copula 'be'</td>
</tr>
<tr>
<td>24</td>
<td>outsider late system morpheme</td>
<td>3rd person 's'</td>
</tr>
<tr>
<td>25</td>
<td>outsider late system morpheme</td>
<td>regular past 'ed'</td>
</tr>
<tr>
<td>26</td>
<td>outsider late system morpheme</td>
<td>auxiliary 'be' 'have/has/had' 'do/does/did'</td>
</tr>
<tr>
<td>27</td>
<td>outsider late system morpheme</td>
<td>clitic (n't), affix (un-, -ness)</td>
</tr>
<tr>
<td>28</td>
<td>outsider late system morpheme</td>
<td>infinitive</td>
</tr>
<tr>
<td>29</td>
<td>total TLU (target-like usage: accuracy) %</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>overall content morpheme TLU</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>overall early system morpheme TLU</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>overall bridge late system morpheme TLU</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>overall outsider late system morpheme TLU</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>overall late system morpheme (bridge + outsider) TLU</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>overall system morpheme (early + bridge + outsider) TLU</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>syntax accuracy %</td>
<td></td>
</tr>
</tbody>
</table>
Derivational morphological forms to be examined are as follows: (1) noun suffixes of 'ist', 'ity', 'ion', and 'ation', (2) verb suffixes of 'ify', 'ize', and 'ate', (3) adjectives 'ous', 'ious', 'ive', and 'al', and (4) derivational pairs of 'person-personal', 'space-spatial', and 'nature-natural'.

The 4-M model (Myers-Scotton, 2002) is also used for accuracy analysis of four types of morphemes to reveal different developmental stages. Examples of morphemes used in this study replicate what Taura (2008) used in his bilingual study (Table 8).

4.4 Syntax

4.4.1 Intrasentential level

As indicators of the length of text, the mean number of words in each T-unit and the mean number of (subordinate) clauses in each T-unit are calculated from the spoken and written data. Other variables we use include (1) syntactically uncommon structures of appositives, elaborated subjects, post modification via prepositional phrases, relative clauses, and non finite verbs, (2) complex and lengthy verb phrases of modal auxiliary verbs, the perfect aspect, and the passive voice, (3) the sequence of a main clause and a preceding subordinate clause (the main clause coming first is more frequent with smaller children), and (4) the type of cohesion devices, that is conjunctions that are subordinating like 'when' and 'although', coordinating like 'and' and 'but', or correlative like 'both' and 'either'.

4.4.2 Intersentential level

Adverbial conjunctions and phrases explored include 'accordingly', 'actually', 'anyway', 'basically', 'before I forget' 'consequently', 'contrastively', 'conversely', 'definitely', 'essentially', 'eventually', 'for example', 'furthermore', 'generally', 'however', 'incidentally' 'literally', 'meanwhile', 'moreover', 'nevertheless', 'normally', 'obviously', 'originally', 'otherwise', 'personally', 'rather', 'similarly', 'that reminds me' 'therefore', 'thus', 'typically', 'unfortunately', and 'usually'.
4.4.3 Tasks

Discourse genre plays a crucial role in syntactic complexity, therefore the type of data collected should be kept constant - in our study the bilingual children produce story-telling (using Mayer's frog series) and casual conversation with the current researchers as spoken data while they undertake a standardized writing test (TOWL-3) designed by Hammill and Larsen (1996) as written data.

4.5 Speaking (conversation and narration)

We employ the following variables: (1) topic maintenance percentage, (2) the number of utterances per topic including the number of words and episodes, (3) the number of smaller episodes embedded in a larger plot, (4) topic shading percentage (development of the ability to stay on topic and to shift gracefully from one topic to another), (5) cohesive devices, that is conjunctions for a smoother flow to express additive ('and', 'also'), temporal ('meanwhile', 'next'), causal ('so', 'therefore'), and adversative ('yet', 'but') meaning, and (6) how often (%) the characters' emotions, thoughts, and feelings are expressed by using emotional words such as 'happy', 'sad', and 'scared'.

Additionally, individual differences are looked at in two respects: (1) whether reading habits influence their story-telling as the elder bilingual child reads a lot more than his sister and (2) whether the male bilingual displays more attempts than the female participant to entertain the interlocutors with jokes.

As for the variables to examine speech fluency, we employ 37 temporal variables and 7 pause distribution variables (Table 5).

4.6 Writing

The numbers of words, subordinate clauses, embeddings, and the passive voice in each writing sample are counted.
The nature of writing requires that explicit instruction, frequent practice opportunity, and feedback from teachers and parents is presented. None of these opportunities are sufficiently provided to our subjects who live in a Japanese dominant community. Thus their developmental course in their writing is explored based on this background.

Keeping the writing task genre constant is a key in analyzing the collected written data reliably, which results in us having to use the same standardized writing test of TOWL-3. The scores of three sections - contextual conventions (CC), contextual language (CL), and story construction (StC) are provided categorized by age along with the total score called "Quotient".

REFERENCES
Bilingual First Language Development of Two Japanese-English Bilingual Siblings: Literature Review on Possible Dependent Variables


Frequency Level Checker Software (http://language.tiu.ac.jp/flc/).


Mahony, D., Singson, M., and Mann, V. (2000). Reading ability and sensitivity to morphological relation. *Reading and Writing*, 12,


Tasks. Denver: National Assessment of Educational Progress.


Weber-Fox, C. and Neville, H.J. (2001). Sensitive periods differentiate processing of open- and closed-class words: An ERP study of


